cmajOPEN

Research

Canadians' views on the use of routinely collected data in health research: a patient-oriented cross-sectional survey

Natalie McCormick PhD, Clayon B. Hamilton PhD, Cheryl L. Koehn BA, Kelly English, Allan Stordy, Linda C. Li PT PhD

Abstract

Background: Little is known about Canadians' knowledge of and level of support for using administrative and other large, routinely collected data for health research, despite the benefits of this type of research to patients, health care systems and society. We sought to benchmark the views of Canadian adults on this topic.

Methods: Researchers and patient leaders of 3 joint and skin disease organizations codeveloped a cross-sectional online survey that was conducted between January and August 2017. The patient partners were engaged as full partners. Recruitment was mainly through the organizations' websites, email and social media. The survey captured respondents' initial perceptions, then (after background information on the topic was provided) elicited their views on the benefits of health research using routinely collected data, data access/privacy concerns, ongoing perceptions and educational needs.

Results: Of the 230 people who consented, 183 (79.6%) started the survey, and 151 (65.6%) completed the survey. Of the 151, 117 (77.5%) were women, 84 (55.6%) were British Columbians, 87 (57.6%) were university graduates, and 101 (66.9%) had a chronic disease. At the beginning of the survey, 119 respondents (78.8%) felt positively about the use of routinely collected data for health research. Respondents identified the ability to study long-term treatment effects and rare events (114 [75.5%]) and large numbers of people (110 [72.8%]) as key benefits. Deidentification of personal information was the top privacy measure (135 [89.4%]), and 101 respondents (66.9%) wanted to learn more about data stewards' granting access to data. On survey completion, more respondents (141 [93.4%]) felt positively about the use of routinely collected data, but only 87 (57.6%) were confident about data security and privacy.

Interpretation: Respondents generally supported the use of deidentified routinely collected data for health research. Although further investigation is needed with more representative samples, our findings suggest that additional education, especially about access and privacy controls, may enhance public support for research endeavours using these data.

Plain language summary: Every day, information is collected about how Canadians use publicly funded health care. The information collected is stored in large electronic data sets. Studying these data can teach us about the quality of our health care, help patients and improve government spending on health care. For example, some data sets can show long-term effects of a given type of treatment. This is not commonly known from single research studies involving small numbers of patients. It is not clear how much Canadians know about these data sets and how researchers can use them. This study surveyed adult Canadians on their views on using such data sets to do health research. The study was done in partnership between university researchers and 3 patients representing national arthritis and skin disease groups. Together, they wrote and tested an online survey, found participants, and interpreted and shared what they learned from the study. Most people who finished the survey agreed with using large data sets to do health research, especially after they read background information on their value. The respondents said the best uses were shown when researchers studied long-term effects in large numbers of people. They felt names and personal information should not be shown. They wished to learn more about how researchers are given access to data sets, and, at the end of survey, they still felt unsure about how privacy and security are ensured. In conclusion, providing information to the public on the use of large sets of health data may increase support for this type of health research.

Analyses of data collected routinely by provincial health ministries and other public bodies[1] can inform health policy-making[2] and advance our knowledge of the burden[3] and risks[4] of diseases, the clinical effectiveness[5] and cost-effectiveness[6] of treatments, and drivers of health care costs.[7] Health researchers in Canada can access a growing array of publicly collected data on the use

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Linda Li, lli@arthritisresearch.ca

***CMAJ Open* 2019. DOI:10.9778/cmajo.20180105**

of health care services, workplace safety, immigration and early childhood development, and link them with other health data including electronic medical records[8] and cancer and perinatal registries. For example, by linking 17 sources of data, Ontario researchers assessed regional variation in cardiovascular event rates[9] and the added value of high-intensity statin therapy for older patients.[10] A crucial advantage of these "real-world" data is their coverage of Canadians of all ages and ethnicities.[11] This makes findings more applicable to the entire Canadian population, including groups who tend not to participate in studies. Despite these benefits, an expert panel of researchers, data custodians, ethicists and managers of health research organizations convened by the Canadian Institutes of Health Research concluded that researchers face many barriers in accessing health and social data in a timely manner.[12] Interestingly, the panel felt many of these barriers stemmed from institutions' being "overly cautious" in their interpretations of privacy legislation and public attitudes about this use of health data.[12] The panel suggested that these barriers could be reduced through greater public engagement. It may be that current data-access policies and practices are stricter than most people would prefer given the societal benefits of this type of research and deidentified nature of the data.

Although a 2016 report by the Office of the Privacy Commissioner of Canada suggested that Canadians have concerns about organizations' collecting and using their personal information,[13] little is known regarding their privacy concerns with the use of publicly collected health data for health research. There is also a dearth of Canadian research on people's understanding of what these data sets contain, what they are used for, who can access them and how they are protected. Prior Canadian studies focused on a single issue (namely, consent preferences for research use of medical records)[14–16] or sampled patients with specific diagnoses.[17–19] In the present study, researchers and patient leaders from 3 national joint and skin disease organizations developed and distributed an online survey to benchmark the views of Canadians about the use of large sets of routinely collected data in health research.

Methods

In this patient-initiated study, we collected data via an online survey among Canadians aged 18 years or more, in English and French, between Jan. 17 and Aug. 15, 2017. The survey focused on administrative data collected by provincial/federal ministries and other public bodies, with access controlled by data stewards,[1] officials who adjudicate researchers' requests for deidentified subsets of these data. We chose a Web-based format so responses could be collected from different areas of Canada and analyzed efficiently. The survey was hosted on the University of British Columbia's FluidSurveys platform (https://it.ubc.ca/services/teaching-learning-tools/survey-tool-fluidsurveys). All collected data remained in Canada, on a secured network server at the researchers' offices at Arthritis Research Canada.

Questionnaire design

Development of the questionnaire was co-led by C.L.K., N.M. and C.B.H., with contributions from all authors. N.M. subsequently modified some questions and content based on her strong background in the acquisition, validation and use of administrative data in health research.[20–25] The draft questionnaire was reviewed and tested by the other patient partners, who provided iterative feedback via email to ensure it could be completed within 15 minutes and its content would be understandable and interesting to lay audiences. Although the survey focused on administrative data, we used the more recognizable (although somewhat less suitable) term "big data" to enhance its accessibility. The questionnaire was formulated in English as a Word document, then translated into French by a professional translator and programmed online. Researchers and patient partners provided additional feedback on the online version, focusing on ease of navigation.

The final questionnaire consisted of 6 sections along with a preamble that introduced respondents to the survey's topic and purpose (Appendix 1, available at www.cmajopen.ca/content/7/2/E203/suppl/DC1). The first section asked about respondents' familiarity with and initial perceptions about the use of big data in health research. In the next 4 sections, respondents were queried about their views on specific topics, including perceived benefits of using these data, types of health research questions that can be answered, and data access and privacy measures. In the final section, respondents were queried further about their perceptions and interest in learning more following completion of the survey. The questionnaire was designed primarily to ascertain respondents' views, but a small amount of educational information was provided since we expected that most respondents would not be familiar with these data sets. For example, when asking which privacy measures were most important, a description was provided of each. Finally, the same question ("In general, how do you feel about the use of big data for health research?") was posed at the start and end of the survey to explore whether respondents' perceptions changed as they completed it.

Recruitment

Respondents were recruited online, mainly through the websites, email newsletters[26] and social media channels of the patient partners' affiliated organizations. These efforts were supplemented by direct emails, social media and word-of-mouth communication between the grant investigators and their colleagues at health research institutes and patient organizations throughout Canada. All of these recruitment strategies were in English and French. Recruitment notices contained a link to the consent page listing the principal investigator, title and purpose of the study, and how and where responses would be stored. Participants were required to review the page, confirm that they were at least 18 years of age and consent to participate before starting the survey. All questions were optional, no personal identifiers were collected, and no incentives were offered.

Patient engagement

The study was conducted at all stages through a partnership with researchers at Arthritis Research Canada; more details are provided in Table 1,[26,28,29] which was guided by the Patient Engagement in Research Description Framework,[29] as used previously.[30] The patient partners, 2 women and 1 man, represent national organizations pertaining to their disease: Arthritis Consumer Experts (C.L.K.), the Arthritis Patient Advisory Board of Arthritis Research Canada (K.E.) and the Canadian Skin Patient Alliance (A.S.). Led by C.L.K., the patient partners are coinvestigators on a multidisciplinary team funded by the Canadian Institutes of Health Research[27] aiming to advance knowledge about the prevention, burden and management of serious complications in chronic inflammatory diseases. Their role includes codeveloping projects to engage the public about the conduct of health research. Having successfully administered an online survey about Canadians' research priorities[31] that informed the grant's development, they sought to increase people's awareness and understanding about using routinely collected data in health research. This need arose through the frequent interactions they had, as leaders of their respective research and advocacy organizations, with patients, investigators, clinicians and the public at large about the Canadian health research landscape.

The patient partners teamed up with a doctoral candidate with expertise in using administrative data for health services research (N.M.),[20–24] a postdoctoral fellow studying patient engagement in research (C.B.H.) and a senior researcher in knowledge translation and implementation science (L.C.L.). During the team's initial meetings, C.L.K. explained how the patient partners aimed to create a series of communiqués to educate the public. As a first step, they wanted to conduct a survey to benchmark people's current understanding, support, concerns and willingness to learn more about this type of research.

Statistical analysis

Only responses by respondents who formally submitted the survey were analyzed; missing responses were permitted for individual questions within these submissions. For each question, we calculated the proportion of respondents who selected each item. Since respondents were often asked to select multiple items from a list, the sum of percentages/frequencies could exceed 100%. Analysis was conducted with SAS software, version 9.5 (SAS Institute) and presented to the team at in-person meetings. The patient partners reviewed and affirmed the results, discussed their possible impact for policy-makers, researchers and the public, and developed a dissemination plan.

Table 1: Outline of partnership between patient partners and researchers in the research process

Stage	Activity
Preparatory	• As coinvestigators on a multidisciplinary team funded by the Canadian Institutes of Health Research,[27] the patient partners (C.L.K., K.E. and A.S.) were allocated funds and approached researchers with their research idea • Researchers and patient partners met to discuss the purpose of the survey and to agree on a preliminary timeline and allocation of tasks and responsibilities
Execution	
Designing study	• Research trainees (N.M. and C.B.H.) and lead patient partner (C.L.K.) codeveloped the initial questionnaire • Patient partners suggested questions based on his/her personal experiences and interactions with people about this type of research • Researchers modified some questions based on their expertise in designing surveys and analyzing health data • Patient partners provided iterative feedback to ensure the survey could be completed within 15 min and the content would be understandable and interesting to lay audiences • Patient partners recommended adding a progress bar (A.S.) and emphasizing that the physician billings data do not contain comments from patients' charts (K.E.)
Recruiting; analyzing and interpreting data	• Patient partners advertised the survey through their personal Twitter feeds and their organizations' websites, email lists and social media channels • K.E. interviewed the first author for her organization's newsletter[26] to promote the survey and inform readers about this type of health research • Researchers promoted the survey to their colleagues at health research institutes and patient organizations throughout Canada • C.B.H. conducted the statistical analysis and presented the results to the team at in-person meetings • Patient partners reviewed and affirmed the results and discussed their possible impact for policy-makers, researchers and the public
Translation	• Research trainees drafted the abstract and manuscript, which the principal investigator (L.C.L.) and patient partners reviewed and edited for critically important content • Researchers and patient partners copresented a poster of the study findings at a major scientific meeting,[28] where it was selected for a poster tour • Patient partners along with researchers produced a lay summary of study findings • Patient partners and researchers are codeveloping materials to educate the public, based on the learning needs and interests identified from the survey • Patient partners and researchers may undertake further studies to assess people's preferences about the terminology to use

Ethics approval

The study received ethics approval from the University of British Columbia's Behavioural Research Ethics Board.

Results

A total of 230 people consented to participate, of whom 151 (65.6%) submitted responses and were included in the analysis. Of the 79 responses not submitted, 47 were from people who consented to participate but never started the survey, and 32 were from people who answered at least the first question but did not complete and submit the entire survey. Most of the surveys submitted (140 [92.7%)] were completed in English. More than three-quarters of the respondents (117 [77.5%]) were women, 71 (47.0%) were aged 50–69 years, and 42 (27.8%) were aged 30–49 years (Table 2). Most lived in British Columbia (84 [55.6%]) or Ontario (39 [25.8%]), over half (87 [57.6%]) had a university degree, and 101 (66.9%) reported having a chronic disease.

Initial knowledge and perceptions

A total of 119 respondents (78.8%) felt positively about the use of routinely collected data for health research, 30 (19.9%) did not know, 1 (0.7%) felt negatively, and 1 (0.7%) declined to answer. Almost all respondents (144 [95.4%]) had heard of the term "electronic health/medical record"; fewer (88 [58.3%]) had heard of the terms "administrative health database" or "administrative data."

Perceived uses and benefits

The ability to study long-term effects and rare events (114 respondents [75.5%]) and large numbers of people (110 [72.8%]) were perceived to be the most important benefits of using routinely collected data (Table 3). Similarly, potentially harmful treatments (97 [64.2%]) and long-term effects/rare events (96 [63.6%]) were the top benefits they wanted to learn more about. Respondents selected long-term harms and benefits of a particular treatment and complications of a particular disease as the most important issues related to health policy and patient care (79–84 [52.3%–55.6%]).

Access and privacy

The need to apply for the use of research data (94 [62.2%]) and the need to obtain approval from university research ethics boards (89 [58.9%]) were the top-ranked data access controls (Table 4). Two-thirds of respondents (101 [66.9%]) wanted to learn more about the role of data stewards in granting access. Deidentifying personal information was selected by the most respondents (135 [89.4%]) as an important measure to enhance information security and protect privacy, followed by mandating researchers to complete privacy training and sign confidentiality agreements (87 [57.6%]).

Next steps

The vast majority of respondents thought the provinces should promote the use of routinely collected data for health research (137 [90.7%]) and were very or somewhat willing to have their deidentified data used by Canadian health researchers (140 [92.7%]). When asked how they felt about the use of routinely collected data for health research at the end of the survey, more felt positively (141 [93.4%]) than at the start. Even still, only 87 respondents (57.6%) were confident about the privacy and security measures in place. The top concern was data access by insurance companies (89 [58.9%]). The potential cost of collecting, storing and overseeing the data was not a major concern (15 [9.9%]). Web sites were the preferred mode for learning more about the use of routinely collected data in Canadian health research (133 [88.1%]).

Table 2: Respondents' characteristics

Characteristic	No. (%) of participants *n* = 151
Sex	
Female	117 (77.5)
Male	32 (21.2)
Declined to answer	2 (1.3)
Responded in French	11 (7.3)
Age group, yr	
18–29	21 (13.9)
30–49	42 (27.8)
50–69	71 (47.0)
70–79	16 (10.6)
≥ 80	1 (0.7)
Education level	
High school or less	10 (6.6)
Some community, technical, trade or vocational college	23 (15.2)
Community college degree/diploma or some university	30 (19.9)
University degree or higher	87 (57.6)
Declined to answer	1 (0.7)
Province of residence*	
British Columbia	84 (55.6)
Alberta	7 (4.6)
Ontario	39 (25.8)
Quebec	13 (8.6)
New Brunswick	3 (2.0)
Nova Scotia	2 (1.3)
Prince Edward Island	1 (0.7)
Declined to answer	2 (1.3)
Chronic disease	
Yes	101 (66.9)
No	44 (29.1)
Declined to answer	6 (4.0)

*No responses were received from Saskatchewan, Manitoba, Newfoundland and Labrador, the Northwest Territories, the Yukon Territory or Nunavut.

Table 3: Responses regarding reasons to use routinely collected data for health research

Area; item	No. (%) of respondents $n = 151$	
	Most important*	Want additional information about†
Benefits of using big data		
Study long-term effects and rare events	114 (75.5)	96 (63.6)
Study large numbers	110 (72.8)	67 (44.4)
Study potentially harmful treatments	76 (50.3)	97 (64.2)
General population comparisons	70 (46.4)	63 (41.7)
More inclusive	65 (43.0)	53 (35.1)
Benefits of using big data from Canada		
Reflective of Canadian health care system	100 (66.2)	–
More inclusive	96 (63.6)	–
Universal prescription medication data	54 (35.8)	–
Reflective of Canadian population	42 (27.8)	–
Issues to study using big data		
Treatment benefits	84 (55.6)	–
Treatment harms	83 (55.0)	–
Disease complications	79 (52.3)	–
Changes in policy or practice	66 (43.7)	–
Quality of care	46 (30.5)	–
Cost-effectiveness	42 (27.8)	–
Risk factors for disease	35 (23.2)	–
Disease incidence and prevalence	15 (9.9)	–

*Respondents could select up to 3 items for benefits of using big data and issues to study using big data, and up to 2 items for benefits of using big data from Canada.
†Respondents could select up to 3 items.

Table 4: Responses regarding data access and privacy and security controls

Area; measure	No. (%) of respondents $n = 151$	
	Most important*	Want additional information about†
Data access controls		
Must apply for data access	94 (62.2)	38 (25.2)
Approval from research ethics board	89 (58.9)	70 (46.4)
Approval from data stewards	77 (51.0)	101 (66.9)
Access data for limited time	31 (20.5)	45 (29.8)
Privacy and security controls		
Data are deidentified	135 (89.4)	–
Privacy training and confidentiality agreement	87 (57.6)	–
Review of research outputs	66 (43.7)	–
Funding agencies cannot access data	54 (35.8)	–
No access outside Canada	53 (35.1)	–

*Respondents could select up to 2 measures for data access controls and up to 3 measures for privacy and security controls.
†Respondents could select up to 2 measures.

Interpretation

By the end of the survey, more than 90% of our respondents felt positively about having publicly collected data available and used for health research, particularly for studying rare health conditions and the long-term consequences of treatments. Consistent with reports from other countries,[32–36] respondents placed high importance on deidentification of data. However, although 93% were at least somewhat willing to have their deidentified information used by health researchers in Canada, far fewer (58%) were confident about the privacy and security procedures in place.

A strength of this study was the active role the patient partners took in interpreting the results. They were particularly interested in how respondents' views changed during the survey: 79% felt positively about the use of routinely collected data at the beginning, compared to 93% at the end. Response bias may have contributed if respondents thought they should feel more positively by the end. However, the change may also have been due to the small amount of educational information provided alongside the survey questions. Although this hypothesis should be tested in future surveys, it is supported by studies conducted in the United Kingdom[37,38] and New Zealand[39] in which participants reported being more comfortable about the use of health data for research after receiving more information. This potentially influential role of education is important given our respondents' desire to learn more about specific topics and their lack of confidence in existing privacy and security procedures. To address these needs, we recommend showcasing Canadian studies in which administrative data were used to assess complications[40–42] and long-term effects[43–46] of medications in patients with chronic diseases. Furthermore, it is essential to provide more information about the role of data stewards in adjudicating data requests and imposing conditions to data access. Prior work suggests that members of the public benefit from hearing the views of the different parties involved and affected by this type of research, including researchers, public health leaders and patients.[38,39,47] As such, the researchers and patients who partnered on this survey are codeveloping educational resources that incorporate patients' perspectives. These should complement the lay summaries, videos and other educational materials available from organizations such as

Population Data BC (https://www.popdata.bc.ca/) and ICES that facilitate research with health data.

Engaging the public should also help in reaching consensus about the terminology used to describe this type of research. As per the patient partners' suggestion, we used the somewhat catchy term big data, although, strictly speaking, structured administrative data sets do not fulfill the characteristics of the big data used in health research,[48] which are typically unstructured. These characteristics include high volume, variety (linking multiple sources such as administrative, survey and electronic medical record data)[48] and velocity (data collected daily or in real time).[48] However, as we believe that most people are not aware of these distinctions, we felt that big data, a mainstream term used to describe this type of health research in Canada[49,50] and elsewhere,[51–53] would resonate better than "administrative" or "publicly/routinely collected" data.

Limitations

Our convenience sample was small, with over half of the respondents residing in BC and only 7% responding in French. Over half were university graduates, although the impact of education level on privacy concerns about health information is unclear,[54] with several studies indicating no differences.[16,18,19] We did not formally assess the face or content validity of our questionnaire, and, although it was translated from English to French by an experienced professional translator, no back-translation[55] or further refinements[56] were done to correct mistranslations. Although the survey was open to all Canadian adults, recruitment was carried out mainly through groups for patients with arthritis or skin disease, and our findings may not reflect the privacy and security concerns of people with other health conditions or the public at large. An earlier survey that the patient partners conducted to inform the grant application was much shorter and received 636 responses over 3 weeks.[31] Although we could not calculate the recruitment/participation rate for the current survey, we suspect that many potential respondents, especially those who found the survey through social media, reviewed the consent page and perceived the survey as too long or too formal. As such, although we intended to recruit a larger, more generalizable sample than those in prior Canadian investigations that limited sampling to patients with specific diagnoses (182–235 respondents each),[17–19] our findings cannot be generalized to the entire Canadian population and must be interpreted cautiously.

Lessons learned from patient involvement

The patient partners' engagement added credibility to the study and its findings. It may have appeared self-serving for the researchers to conduct this survey on their own and report a favourable level of support among people in Canada for the type of research they are personally involved with. The research trainees gained valuable experience in collaborating with patient partners and developed a better understanding of partners' roles, such as researchers' role in maintaining scientific rigor and patient partners' role in driving study accessibility. In addition, through the guidance of the patient partners and mentoring of an experienced researcher, the team negotiated strategies to facilitate participation of patient partners over the course of this study. Notably, the team gained experience in discussing authorship. In turn, the patient partners learned more about the research process, including how researchers go about accessing administrative data, and copresented the findings at the American College of Rheumatology annual meeting.[28]

Conclusion

As new sources of publicly and patient-collected information, including electronic medical records, biospecimens/genomic data, wearable devices and mobile health apps, become available for linkage and analysis, it is important to engage with Canadians about health research using this type of data and their preferences for data-access and data-sharing policies. Although the vast majority of our respondents felt positively about the use of publicly collected data, many lacked confidence in the access and privacy controls. Our findings warrant further investigation with more population-based sampling strategies so that the views of people of different ages, races/ethnicities, socioeconomic backgrounds, languages and health status are better represented. The extent to which wording (i.e., "personal health information" v. "administrative data") and level of detail in the survey (i.e., whether to mention that consent is not sought for this type of research) affect respondents' attitudes should also be investigated, along with any changes in concern about insurance companies' accessing health data since the passage of Canada's Genetic Non-Discrimination Act,[57] in mid-2017. For now, this dynamic partnership between researchers and patients was an important first step in understanding the views of Canadians on the use of large sets of routinely collected health information for health research.

References

1. Glossary. PopulationData BC. Available: https://www.popdata.bc.ca/BCDataDeliberation/glossary (accessed 2018 June 12).
2. Lavergne MR, Law MR, Peterson S, et al. A population-based analysis of incentive payments to primary care physicians for the care of patients with complex disease. *CMAJ* 2016;188:E375-83.
3. Benchimol EI, Bernstein CN, Bitton A, et al. Trends in epidemiology of pediatric inflammatory bowel disease in Canada: distributed network analysis of multiple population-based provincial health administrative databases. *Am J Gastroenterol* 2017;112:1120-34.
4. Lega IC, Wilton AS, Austin PC, et al. The temporal relationship between diabetes and cancer: a population-based study. *Cancer* 2016;122:2731-8.
5. Rawson NSB, Downey W, Maxwell CJ, et al. 25 years of pharmacoepidemiologic innovation: the Saskatchewan health administrative databases. *J Popul Ther Clin Pharmacol* 2011;18:e245-9.
6. Khor S, Beca J, Krahn M, et al. Real-world costs and cost-effectiveness of rituximab for diffuse large B-cell lymphoma patients: a population-based analysis. *BMC Cancer* 2014;14:586.
7. Fitzpatrick T, Rosella LC, Calzavara A, et al. Looking beyond income and education: socioeconomic status gradients among future high-cost users of health care. *Am J Prev Med* 2015;49:161-71.
8. Tu K, Mitiku TF, Ivers NM, et al. Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD). *Am J Manag Care* 2014;20:e15-21.
9. Tu JV, Chu A, Maclagan L, et al. Regional variations in ambulatory care and incidence of cardiovascular events. *CMAJ* 2017;189:E494-501.
10. Alter DA, Tu JV, Koh M, et al. Projected real-world effectiveness of using aggressive low-density lipoprotein cholesterol targets among elderly statin users following acute coronary syndromes in Canada. *J Am Heart Assoc* 2018;7:e007535.
11. Lougheed T. Rethinking barriers to big data. *CMAJ* 2015;187:E1.
12. Accessing health and health-related data in Canada. Ottawa: Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, Council of Canadian Academies; 2015. Available: www.scienceadvice.ca/reports/accessing-health-and-health-related-data-in-canada/ (accessed 2019 Mar. 6).

13. 2016 survey of Canadians on privacy: final report. Gatineau (QC): Office of the Privacy Commissioner of Canada; 2016. Available: https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2016/por_2016_12/ (accessed 2018 Sept. 16).
14. Nair K, Willison D, Holbrook A, et al. Patients' consent preferences regarding the use of their health information for research purposes: a qualitative study. *J Health Serv Res Policy* 2004;9:22-7.
15. Willison DJ, Swinton M, Schwartz L, et al. Alternatives to project-specific consent for access to personal information for health research: insights from a public dialogue. *BMC Med Ethics* 2008;9:18.
16. Willison DJ, Schwartz L, Abelson J, et al. Alternatives to project-specific consent for access to personal information for health research: What is the opinion of the Canadian public? *J Am Med Inform Assoc* 2007;14:706-12.
17. Page SA, Manhas KP, Muruve DA. A survey of patient perspectives on the research use of health information and biospecimens. *BMC Med Ethics* 2016;17:48.
18. Terry AL, Chesworth BM, Bourne RB, et al. Joint replacement recipients' views about health information privacy. *Health Expect* 2015;18:1519-29.
19. Page SA, Mitchell I. Patients' opinions on privacy, consent and the disclosure of health information for medical research. *Chronic Dis Can* 2006;27:60-7.
20. McCormick N, Reimer K, Famouri A, et al. Filling the gaps in SARDs research: collection and linkage of administrative health data and self-reported survey data for a general population-based cohort of individuals with and without diagnoses of systemic autoimmune rheumatic disease (SARDs) from British Columbia, Canada. *BMJ Open* 2017;7:e013977.
21. McCormick N, Marra CA, Sadatsafavi M, et al. Incremental direct medical costs of systemic lupus erythematosus patients in the years preceding diagnosis: a general population-based study. *Lupus* 2018;27:1247-58.
22. McCormick N, Lacaille D, Bhole V, et al. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One* 2014;9:e92286.
23. McCormick N, Lacaille D, Bhole V, et al. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. *PLoS One* 2014;9:e104519.
24. McCormick N, Bhole V, Lacaille D, et al. Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. *PLoS One* 2015;10:e0135834.
25. Rai SK, Aviña-Zubieta JA, McCormick N, et al. The rising prevalence and incidence of gout in British Columbia, Canada: population-based trends from 2000 to 2012. *Semin Arthritis Rheum* 2017;46:451-6.
26. English K. Big Data: accessing existing health databases to advance health research. Richmond (BC): 2017. Arthritis Research Canada. Available: https://www.arthritisresearch.ca/big-data-accessing-existing-health-databases/ (accessed 2018 Dec. 20).
27. Esdaile J, Abrahamowicz M, Aviña-Zubieta A, et al. PRECISION: Preventing Complications from Inflammatory Skin, Joint and Bowel Conditions. Richmond (BC): Arthritis Research Canada. Available: www.arthritisresearch.ca/precision (accessed 2018 Apr. 21).
28. McCormick N, Hamilton C, Koehn C, et al. Canadians' views about using big data in health research from a national online survey: a partnership of patient-consumers and researchers [abstract]. *Arthritis Rheumatol* 2018;70(Suppl 10).
29. Hamilton CB, Leese JC, Hoens AM, et al. Framework for advancing the reporting of patient engagement in rheumatology research projects. *Curr Rheumatol Rep* 2017;19:38.
30. Leese J, Macdonald G, Kerr S, et al. 'Adding another spinning plate to an already busy life.' Benefits and risks in patient partner–researcher relationships: a qualitative study of patient partners' experiences in a Canadian health research setting. *BMJ Open* 2018;8:e022154.
31. Koehn C, Attara G, Stordy A, et al. Consumer informing research: a survey of Canadians' views and research priorities in chronic inflammatory diseases [abstract]. *J Rheumatol* 2014;41:1465.
32. Xafis V. The acceptability of conducting data linkage research without obtaining consent: lay people's views and justifications. *BMC Med Ethics* 2015;16:79.
33. Buckley BS, Murphy AW, MacFarlane AE. Public attitudes to the use in research of personal health information from general practitioners' records: a survey of the Irish general public. *J Med Ethics* 2011;37:50-5.
34. Luchenski SA, Reed JE, Marston C, et al. Patient and public views on electronic health records and their uses in the United Kingdom: cross-sectional survey. *J Med Internet Res* 2013;15:e160.
35. Zalin A, Papoutsi C, Shotliff K, et al. The use of information for diabetes research and care: patient views in West London. *Pract Diabetes* 2016;33:81-86a.
36. Whiddett R, Hunter I, Engelbrecht J, et al. Patients' attitudes towards sharing their health information. *Int J Med Inform* 2006;75:530-41.
37. Hill EM, Turner EL, Martin RM, et al. "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. *BMC Med Res* 2013;3:72.
38. Tully MP, Bozentko K, Clement S, et al. Investigating the extent to which patients should control access to patient records for research: a deliberative process using citizens' juries. *J Med Internet Res* 2018;20:e112.
39. Parkin L, Paul C. Public good, personal privacy: a citizens' deliberation about using medical information for pharmacoepidemiological research. *J Epidemiol Community Health* 2011;65:150-6.
40. Feldman DE, Vinet É, Bérard A, et al. Heart disease, hypertension, gestational diabetes mellitus, and preeclampsia/eclampsia in mothers with juvenile arthritis: a nested case–control study: maternal outcomes in juvenile arthritis. *Arthritis Care Res (Hoboken)* 2017;69:306-9.
41. Kuenzig ME, Barnabe C, Seow CH, et al. Asthma is associated with subsequent development of inflammatory bowel disease: a population-based case–control study. *Clin Gastroenterol Hepatol* 2017;15:1405-12.e3.
42. Solomon DH, Love TJ, Canning C, et al. Risk of diabetes among patients with rheumatoid arthritis, psoriatic arthritis and psoriasis. *Ann Rheum Dis* 2010;69:2114-7.
43. Raymakers AJN, Sadatsafavi M, Sin DD, et al. The impact of statin drug use on all-cause mortality in patients with COPD. *Chest* 2017;152:486-93.
44. CAN-AIM; Moura CS, Abrahamowicz M, Beauchamp ME, et al. Early medication use in new-onset rheumatoid arthritis may delay joint replacement: results of a large population-based study. *Arthritis Res Ther* 2015;17:197.
45. Dixon WG, Abrahamowicz M, Beauchamp ME, et al. Immediate and delayed impact of oral glucocorticoid therapy on risk of serious infection in older patients with rheumatoid arthritis: a nested case–control analysis. *Ann Rheum Dis* 2012;71:1128-33.
46. Shah M, Avgil Tsadok M, Jackevicius CA, et al. Warfarin use and the risk for stroke and bleeding in patients with atrial fibrillation undergoing dialysis. *Circulation* 2014;129:1196-203.
47. Douglas A, Ward HJT, Bhopal R, et al. Is the linkage of census and health data justified? Views from a public panel of the Scottish Health and Ethnicity Linkage study. *J Public Health (Oxf)* 2018;40:435-40.
48. Fuller D, Buote R, Stanley K. A glossary for big data in population and public health: discussion and commentary on terminology and research methods. *J Epidemiol Community Health* 2017;71:1113-17.
49. Project Big Life [homepage]. Available: https://www.projectbiglife.ca/ (accessed 2019 Jan. 3).
50. Tu JV, Chu A, Donovan LR, et al. The Cardiovascular Health in Ambulatory Care Research Team (CANHEART): using big data to measure and improve cardiovascular health and healthcare services. *Circ Cardiovasc Qual Outcomes* 2015;8:204-12.
51. Kashef Z. Big data analysis accurately predicts patient survival from heart failure. *YaleNews* [New Haven (CT)] 2018 Apr. 12. Available: https://news.yale.edu/2018/04/12/big-data-analysis-accurately-predicts-patient-survival-heart-failure (accessed 2019 Jan. 3).
52. Big data initiative set to make impact on health of society [news release]. Farr Institute of Health Informatic Research 2016 Feb. 29. Available: http://farrinstitute.org/news/big-data-initiative-set-to-make-impact-on-health-of-socie (accessed 2019 Jan. 3).
53. van Staa TP, Goldacre B, Buchan I, et al. Big health data: the need to earn public trust. *BMJ* 2016;354:i3636.
54. Kho ME, Duffett M, Willison DJ, et al. Written informed consent and selection bias in observational studies using medical records: systematic review. *BMJ* 2009;338:b866.
55. Brislin RW. Back-translation for cross-cultural research. *J Cross Cult Psychol* 1970;1:185-216.
56. Santos HPO, Black AM, Sandelowski M. Timing of translation in cross-language qualitative research. *Qual Health Res* 2015;25:134-44.
57. *Genetic Non-Discrimination Act* (S.C. 2017, c. 3). Available: http://laws.justice.gc.ca/eng/AnnualStatutes/2017_3/FullText.html (accessed 2018 Sept. 14).

Affiliations: Arthritis Research Canada (McCormick, Hamilton, Li), Richmond, BC; Department of Physical Therapy (Hamilton, Li), University of British Columbia; Arthritis Consumer Experts (Koehn), Vancouver, BC; Arthritis Patients Advisory Board (English), Arthritis Research Canada, Richmond, BC; Canadian Skin Patient Alliance (Stordy), Calgary, Alta.

Contributors: Clayon Hamilton, Cheryl Koehn, Kelly English, Allan Stordy and Linda Li conceived the study. Clayon Hamilton analyzed the data. Natalie McCormick and Clayon Hamilton drafted the manuscript. All of the authors contributed substantially to the study design and interpretation of data, critically revised the manuscript for important intellectual content, gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

Funding: This study was funded by Team Grant THC-316595 from the Canadian Institutes of Health Research (CIHR). Natalie McCormick was funded by a Doctoral Research Award from CIHR and now holds a CIHR Fellowship Award. Clayon Hamilton is funded by a Michael Smith Foundation for Health Research Trainee Award in Health Services Research and a CIHR Health System Impact Fellowship. Linda Li is supported by the Canada Research Chair Program, the Harold Robinson/Arthritis Society Chair in Arthritic Diseases and the Michael Smith Foundation for Health Research Scholar Award.

Acknowledgements: The authors thank all those who participated in the survey. They also thank Lynn Nowoselski, who coordinated the project.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/7/2/E203/suppl/DC1.